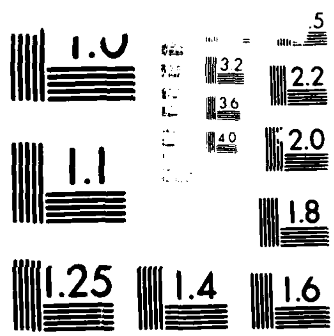
AD-A168 503 SOME THEORETICAL PROPERTIES OF AN AUGMENTED LAGRANGIAN MERIT FUNCTION(U) STANFORD UNIV CA SYSTEMS OPTIMIZATION LAB P E GILL ET AL. APR 86 SOL-86-6 N00014-85-C-0343 1/1

UNCLASSIFIED F/G 12/2 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
3.2
3.6
4.0

Systems
Optimization
Laboratory

AD-A168 503

# SOME THEORETICAL PROPERTIES OF AN AUGMENTED LAGRANGIAN MERIT FUNCTION

by

Philip E. Gill, Walter Murray,
Michael A. Saunders and Margaret H. Wright

TECHNICAL REPORT SOL 86-6

April 1986





Department of Operations Research
Stanford University
Stanford, CA 94305

12

**SYSTEMS OPTIMIZATION LABORATORY**
**DEPARTMENT OF OPERATIONS RESEARCH**
**STANFORD UNIVERSITY**
**STANFORD, CALIFORNIA 94305-4022**

# SOME THEORETICAL PROPERTIES OF AN AUGMENTED LAGRANGIAN MERIT FUNCTION

**by**

**Philip E. Gill, Walter Murray,**
**Michael A. Saunders and Margaret H. Wright**

**TECHNICAL REPORT SOL 86-6**

**April 1986**

Research and reproduction of this report were partially supported by the U.S. Department of Energy Contract DE-AA03-76SF00326, PA# DE-AS03-76ER72018; National Science Foundation Grants DCR-8413211 and ECS-8312142; Office of Naval Research Contract N00014-85-K-0343; and Army Research Office Contract DAAG29-84-K-0156.

# SOME THEORETICAL PROPERTIES OF AN AUGMENTED LAGRANGIAN MERIT FUNCTION



Philip E. Gill, Walter Murray,
Michael A. Saunders and Margaret H. Wright
Systems Optimization Laboratory
Department of Operations Research
Stanford University
Stanford, California 94305

Technical Report SOL 86-6
April 1986



## ABSTRACT

Sequential quadratic programming (SQP) methods for nonlinearly constrained optimization typically use a *merit function* to enforce convergence from an arbitrary starting point. We define a *smooth* augmented Lagrangian merit function in which the Lagrange multiplier estimate is treated as a separate variable, and inequality constraints are handled by means of non-negative slack variables that are included in the linesearch. Global convergence is proved for an SQP algorithm that uses this merit function. We also prove that steps of unity are accepted in a neighborhood of the solution when this merit function is used in a suitable superlinearly convergent algorithm. Finally, a selection of numerical results is presented to illustrate the performance of the associated SQP method.

The material contained in this report is based upon research supported by the U.S. Department of Energy Contract DE-AA03-76SF00326, PA No. DE-AS03-76ER72018; National Science Foundation Grants DCR-8413211 and ECS-8312142; the Office of Naval Research Contract N00014-85-K-0343; and the U.S. Army Research Office Contract DAAG29-84-K-0156.

## 1. Sequential Quadratic Programming Methods

Sequential quadratic programming (SQP) methods are widely considered the most effective general techniques for solving optimization problems with nonlinear constraints. (For a survey of results and references on SQP methods, see Powell, 1983.) One of the major issues of interest in recent research on SQP methods has been the choice of *merit function*— the measure of progress at each iteration. This paper describes some properties of a theoretical SQP algorithm (NPSQP) that uses a *smooth* augmented Lagrangian merit function. NPSQP is a simplified version of an SQP algorithm that has been implemented as the Fortran code NPSOL (Gill *et al.*, 1985, 1986a, b). (The main simplifications involve the form of the problem and strengthened assumptions.)

For ease of presentation, we assume that all the constraints are nonlinear *inequalities*. (The theory applies in a straightforward fashion to equality constraints.) The problem to be solved is thus:

$$\text{NP} \qquad \begin{array}{ll} \underset{x\in\Re^n}{\text{minimize}} & F(x) \\ \text{subject to} & c_i(x) \geq 0, \quad i = 1,\ldots,m, \end{array}$$

where $F$ and $\{c_i\}$ are twice-continuously differentiable. Let $g(x)$ denote the gradient of $F(x)$, and $A(x)$ denote the Jacobian matrix of the constraint vector $c(x)$. The solution of NP will be denoted by $x^*$.

We assume that the first-order Kuhn-Tucker conditions hold (with strict complementarity) at $x^*$, i.e., that there exists a Lagrange multiplier vector $\lambda^*$ such that

$$g(x^*) = A(x^*)^T\lambda^*; \tag{1.1a}$$

$$c(x^*)^T\lambda^* = 0; \tag{1.1b}$$

$$\lambda_i^* > 0 \quad \text{if} \quad c_i(x^*) = 0. \tag{1.1c}$$

(For a detailed discussion of optimality conditions, see, for example, Fiacco and McCormick, 1968, and Powell, 1974.)

At the $k$-th iteration, the new iterate $x_{k+1}$ is defined as

$$x_{k+1} = x_k + \alpha_k p_k, \tag{1.2}$$

where $x_k$ is the current iterate, $p_k$ is an $n$-vector (the *search direction*), and $\alpha_k$ is a non-negative step length $(0 < \alpha_k \leq 1)$. For simplicity of notation, we henceforth suppress the subscript $k$, which will be implicit on unbarred quantities. A barred quantity denotes one evaluated at iteration $k + 1$.

The central feature of an SQP method is that the search direction $p$ in (1.2) is the solution of a quadratic programming subproblem whose objective function approximates the Lagrangian function and whose constraints are linearizations of the nonlinear constraints. The usual definition of the QP subproblem is the following:

$$\underset{p\in\Re^n}{\text{minimize}} \quad g^Tp + \tfrac{1}{2}p^THp \tag{1.3a}$$

$$\text{subject to} \quad Ap \geq -c, \tag{1.3b}$$

where $g$, $c$ and $A$ denote the relevant quantities evaluated at $x$. The matrix $H$ is a symmetric positive-definite quasi-Newton approximation to the Hessian of the Lagrangian function. The Lagrange multiplier vector $\mu$ of (1.3) satisfies

$$g + Hp = A^T\mu; \tag{1.4a}$$

$$\mu^T(Ap + c) = 0; \tag{1.4b}$$

$$\mu \geq 0. \tag{1.4c}$$

The remainder of this paper is organized as follows. Section 2 gives some background on the role of merit functions within SQP methods, and introduces the augmented Lagrangian merit function used in NPSQP. In Section 3 we state the assumptions about the problem and present the algorithm. Global convergence results are given in Section 4. Section 5 shows that the chosen merit function will not impede superlinear convergence. In Section 6 we present a selection of numerical results to indicate the robustness of the associated SQP method NPSOL.

## 2. Background on Merit Functions

### 2.1. Introduction

Many authors have studied the choice of steplength in (1.2). Usually, $\alpha$ is chosen by a linesearch procedure to ensure a "sufficient decrease" (Ortega and Rheinboldt, 1970) in a *merit function* that combines the objective and constraint functions in some way. A popular merit function for several years (Han, 1976, 1977; Powell, 1977) has been the $\ell_1$ penalty function (Pietrzykowski, 1969):

$$P(x,\rho) = F(x) + \rho \sum \max\bigl(0, -c_i(x)\bigr). \tag{2.1}$$

This merit function has the property that, for $\rho$ sufficiently large, $x^*$ is an unconstrained minimum of $P(x,\rho)$. In addition, $\rho$ can always be chosen so that the SQP search direction $p$ is a descent direction for $P(x,\rho)$. However, Maratos (1978) observed that requiring a decrease in $P(x,\rho)$ at every iteration could lead to the inhibition of superlinear convergence (see Chamberlain *et al.*, 1982, for a procedure designed to avoid this difficulty). Furthermore, $P(x,\rho)$ is not differentiable at the solution, and linesearch techniques based on smooth polynomial interpolation are thus not applicable.

An alternative merit function that has recently received attention is the *augmented Lagrangian function*, whose development we now review. If all the constraints of NP are equalities, the associated augmented Lagrangian function is:

$$L(x,\lambda,\rho) \equiv F(x) - \lambda^T c(x) + \tfrac{1}{2}\rho c(x)^T c(x), \tag{2.2}$$

where $\lambda$ is a multiplier estimate and $\rho$ is a non-negative penalty parameter. Augmented Lagrangian functions were first introduced by Hestenes (1969) and Powell (1969) as a means of

creating a sequence of *unconstrained subproblems* for the equality-constraint case. A crucial property of (2.2) is that there exists a *finite* $\hat{\rho}$ such that for all $\rho \geq \hat{\rho}$, $x^*$ is an unconstrained minimum of (2.2) when $\lambda = \lambda^*$ (see, e.g., Fletcher, 1974). The use of (2.2) as a merit function within an SQP method was suggested by Wright (1976) and Schittkowski (1981).

### 2.2. The Lagrange multiplier estimate

When (2.2) is used as a merit function, it is not obvious—even in the equality-constraint case how the multiplier estimate $\lambda$ should be defined at each iteration.

Most SQP methods (e.g., Han, 1976; Powell, 1977) define the approximate Hessian of the Lagrangian function using the QP multiplier $\mu$ (cf. (1.4)), which can be interpreted as the "latest" (and presumably "best") multiplier estimate, and requires no additional computation. However, using $\mu$ as the multiplier estimate in (2.2) has the effect of *redefining* the merit function at every iteration. Thus, since there is no monotonicity property with respect to a single function, difficulties may arise in proving global convergence of the algorithm.

Powell and Yuan (1984) have recently studied an augmented Lagrangian merit function for the equality-constraint case in which $\lambda$ in (2.2) is defined as the *least-squares multiplier estimate*, and hence is treated as a function of $x$ rather than as a separate variable. (An augmented Lagrangian function of this type was first introduced and analyzed as an exact penalty function by Fletcher, 1970.) Powell and Yuan (1984) prove several global and local convergence properties for this merit function.

Other smooth merit functions have been considered by Dixon (1979), DiPillo and Grippo (1979), Bartholomew-Biggs (1983, 1985), and Boggs and Tolle (1984, 1985) (the latter only for the equality-constraint case).

An approach that makes an alternative use of the QP multiplier estimate is to treat the elements of $\lambda$ as *additional variables* (rather than to reset $\lambda$ at every iteration). Thus, a "search direction" $\xi$ is defined for the multiplier estimate, and the linesearch is performed with respect to both $x$ and $\lambda$. This idea was suggested by Tapia (1977) in the context of an unconstrained subproblem and by Schittkowski (1981) within an SQP method. This approach will also be taken in NPSQP.

### 2.3. Treatment of inequality constraints

When defining a merit function for a problem with *inequality* constraints, it is necessary to identify which constraints are "active". The $\ell_1$ merit function (2.1) includes only the *violated* constraints. The original formulation of an augmented Lagrangian function for inequality constraints is due to Rockafellar (1973):

$$L(x,\lambda,\rho) = F(x) - \lambda^T c_+(x) + \tfrac{1}{2}\rho c_+(x)^T c_+(x), \tag{2.3}$$

where

$$(c_+)_i = \begin{cases} c_i & \text{if } \rho c_i \leq \lambda_i; \\ \lambda_i/\rho & \text{otherwise.} \end{cases}$$

Some global results are given for this merit function in Schittkowski (1981, 1983).

A disadvantage of (2.3) as a merit function is that discontinuities in the first derivative may cause difficulties for linesearch techniques based on polynomial interpolation. Therefore, to retain smoothness during the linesearch even with inequality constraints, we augment the variables $(x, \lambda)$ by a set of *slack variables* that are used *only in the line search*. At the $k$-th major iteration, a vector triple

$$y = \begin{pmatrix} p \\ \xi \\ q \end{pmatrix} \tag{2.4}$$

is computed that serves as a direction of search for the variables $(x, \lambda, s)$. The new values are defined by

$$\begin{pmatrix} \bar{x} \\ \bar{\lambda} \\ \bar{s} \end{pmatrix} = \begin{pmatrix} x \\ \lambda \\ s \end{pmatrix} + \alpha \begin{pmatrix} p \\ \xi \\ q \end{pmatrix}, \tag{2.5}$$

where the vectors $p$, $\xi$ and $q$ are found from the QP subproblem (1.3), as described below.

In our algorithm (as in Schittkowski, 1981), $\xi$ is defined as

$$\xi \equiv \mu - \lambda, \tag{2.6}$$

so that if $\alpha = 1$, $\bar{\lambda} = \mu$. We take $\lambda_0$ as $\mu_0$ (the QP multipliers at $x_0$), so that $\lambda_1 = \mu_0$ regardless of $\alpha_0$.

The definition of $s$ and $q$ can be interpreted in terms of an idea originally given by Rockafellar (1973) in the derivation of (2.3). In our algorithm, the augmented Lagrangian function includes a set of non-negative slack variables:

$$L(x, \lambda, s, \rho) = F(x) - \lambda^T\big(c(x) - s\big) + \tfrac{1}{2}\rho\big(c(x) - s\big)^T\big(c(x) - s\big), \quad \text{with} \quad s \geq 0. \tag{2.7}$$

The vector $s$ at the beginning of iteration $k$ is taken as

$$s_i = \begin{cases} \max(0, c_i) & \text{if } \rho = 0; \\ \max(0, c_i - \lambda_i/\rho) & \text{otherwise.} \end{cases} \tag{2.8}$$

(For given values of $x$, $\lambda$ and $\rho$, the vector $s$ defined by (2.8) yields the value of $L$ minimized with respect to the slack variables alone, subject to the non-negativity restriction $s \geq 0$.) The vector $q$ in (2.5) is then defined by

$$Ap - q = -(c - s), \tag{2.9a}$$

so that

$$Ap + c = s + q. \tag{2.9b}$$

We see from (2.9) that $s + q$ is simply the residual of the inequality constraints from the QP (1.3). Therefore, it follows from (1.3) and (1.4) that

$$s + q \geq 0 \quad \text{and} \quad \mu^T(s + q) = 0. \tag{2.10}$$

### 2.4. Choice of the penalty parameter

Finally, we consider the definition of the penalty parameter in (2.7). Our numerical experiments have suggested strongly that efficient performance is linked to keeping the penalty parameter as small as possible, subject to satisfying the conditions needed for convergence. Hence, our strategy is to maintain a "current" $\rho$ that is increased only when necessary to satisfy a condition that assures global convergence.

Several authors have stated that the need to choose a penalty parameter adds an arbitrary element to an SQP algorithm, or leads to difficulties in implementation. On the contrary, we shall see (Lemma 4.3) that in NPSQP *the penalty parameter is based directly on the condition needed for global convergence*, and hence should not be considered as arbitrary or heuristic.

## 3. Statement of Algorithm

We make the following assumptions:

(i) $x$ and $x + p$ lie in a closed, bounded region $\Omega$ of $\Re^n$ for all $k$;

(ii) $F$, $\{c_i\}$, and their first and second derivatives are uniformly bounded in norm in $\Omega$;

(iii) $H$ is positive definite, with bounded condition number, and smallest eigenvalue uniformly bounded away from zero, i.e., there exists $\gamma > 0$ such that, for all $k$,

$$p^T H p \geq \gamma \|p\|^2;$$

(iv) $\|\mu\|$ is uniformly bounded at every iterate.

(v) The quadratic program (1.3) always has a solution.

The following notation will be used in the remainder of this section. Given $x$, $\lambda$, $s$, $p$, $\xi$, $q$ and $\rho$, we let $\phi(\alpha)$ denote $L(x + \alpha p, \lambda + \alpha\xi, s + \alpha q, \rho)$, i.e., the merit function as a function of the steplength. The derivative of $\phi$ with respect to $\alpha$ will be denoted by $\phi'$. Let $v$ denote the "extended" iterate $(x, \lambda, s)$, and let $y$ denote the associated search direction $(p, \xi, q)$. For brevity, we sometimes use the notation "$(\alpha)$" to denote evaluation at $v + \alpha y$.

The steps of each iteration of algorithm NPSQP are:

1. Solve (1.3) for $p$. If $p = 0$, set $\lambda = \mu$ and terminate. Otherwise, define $\xi = \mu - \lambda$.
2. Compute $s$ from (2.8). Find $\rho$ such that $\phi'(0) \leq -\frac{1}{2} p^T H p$ (see Lemma 4.3, below).
3. Compute the steplength $\alpha$, as follows. If

$$\phi(1) - \phi(0) \leq \sigma\phi'(0) \tag{3.1a}$$

and

$$\phi'(1) \leq \eta\phi'(0) \quad \text{or} \quad |\phi'(1)| \leq -\eta\phi'(0), \tag{3.1b}$$

where $0 < \sigma \leq \eta < \frac{1}{2}$, set $\alpha = 1$. Otherwise, use safeguarded cubic interpolation (see, e.g., Gill and Murray, 1974), to find an $\alpha \in (0, 1)$ such that

$$\phi(\alpha) - \phi(0) \leq \sigma\alpha\phi'(0) \tag{3.2a}$$

and

$$|\phi'(\alpha)| \le -\eta\phi'(0). \tag{3.2b}$$

4. Update $H$ so that (iii) is satisfied (e.g., using a suitably modified BFGS update; see Gill *et al.*, 1986b, for the update used in NPSOL).
5. Update $x$ and $\lambda$ using (2.5).

## 4. Global Convergence Results

In order to prove global convergence, we first prove a set of lemmas that establish various properties of the algorithm.

**Lemma 4.1.** *Using Algorithm NPSQP, $\|p\| = 0$ if and only if $x$ is a Kuhn-Tucker point of NP. Furthermore, for $x$ sufficiently close to $x^*$, the active set of the QP (1.3) is the same as the active set of nonlinear constraints for NP.*

**Proof.** See Robinson (1974). ∎

**Lemma 4.2.** *For all $k \ge 1$,*

$$\|\lambda_k\| \le \max_{1\le i\le k-1} \|\mu_i\|,$$

*and hence $\|\lambda_k\|$ is bounded for all $k$. In addition, $\|\xi_k\|$ is uniformly bounded for all $k$.*

**Proof.** By definition,

$$\begin{aligned}\lambda_1 &= \mu_0;\\ \lambda_{k+1} &= \lambda_k + \alpha_k(\mu_k - \lambda_k), \quad k \ge 1.\end{aligned} \tag{4.1}$$

The proof is by induction. The result holds for $\lambda_1$ because of assumption (iv) about the boundedness of $\|\mu_k\|$ for all $k$. Assume that the lemma holds for $\lambda_k$. From (4.1) and norm inequalities, we have

$$\|\lambda_{k+1}\| \le \alpha_k\|\mu_k\| + (1-\alpha_k)\|\lambda_k\|.$$

Since $0 < \alpha \le 1$, applying the inductive hypothesis gives

$$\|\lambda_{k+1}\| \le \max\|\mu_i\|, \quad i = 1, \ldots, k,$$

which gives the first desired result.

The boundedness of $\|\xi_k\|$ follows immediately from its definition (2.6), assumption (iv), and the first result of this lemma. ∎

The next lemma establishes the existence of a non-negative penalty parameter such that the projected gradient of the merit function at each iterate satisfies a condition associated with global convergence.

**Lemma 4.3.** *There exists $\hat{\rho} \ge 0$ such that*

$$\phi'(0, \rho) \le -\tfrac{1}{2} p^T H p \tag{4.2}$$

*for all $\rho \ge \hat{\rho}$.*

**Proof.** The gradient of $L$ with respect to $x$, $\lambda$ and $s$ is given by

$$\nabla L(x, \lambda, s) \equiv \begin{pmatrix} g(x) - A(x)^T \lambda + \rho A(x)^T \big(c(x) - s\big) \\ -\big(c(x) - s\big) \\ \lambda - \rho\big(c(x) - s\big) \end{pmatrix}, \tag{4.3}$$

and it follows that $\phi'(0)$ is given by

$$\phi'(0) = p^T g - p^T A^T \lambda + \rho p^T A^T (c - s) - (c - s)^T \xi + \lambda^T q - \rho q^T (c - s), \tag{4.4}$$

where $g$, $A$, and $c$ are evaluated at $x$.

Multiplying (1.4*a*) by $p^T$ gives

$$g^T p = p^T A^T \mu - p^T H p. \tag{4.5}$$

Substituting (2.6), (2.9*a*) and (4.5) in (4.4), we obtain

$$\phi'(0) = -p^T H p + q^T \mu - 2(c - s)^T \xi - \rho (c - s)^T (c - s). \tag{4.6}$$

Substituting (4.6) in the desired inequality (4.2) and re-arranging, we obtain as the condition to be satisfied

$$q^T \mu - 2(c - s)^T \xi - \rho (c - s)^T (c - s) \le \tfrac{1}{2} p^T H p. \tag{4.7}$$

The complementarity conditions (1.4) and definition (2.9) imply that $q^T \mu \le 0$. Hence, if $\frac{1}{2} p^T H p > -2(c - s)^T \xi$, then (4.7) holds for *all* non-negative $\rho$, and $\hat{\rho}$ may be taken as zero. (Note that this applies when $c - s$ is zero.) The determination of $\hat{\rho}$ is non-trivial only if

$$\tfrac{1}{2} p^T H p \le -2(c - s)^T \xi. \tag{4.8}$$

Re-arranging (4.7), we see that $\rho$ must satisfy

$$\rho (c - s)^T (c - s) \ge q^T \mu - \tfrac{1}{2} p^T H p - 2(c - s)^T \xi.$$

A value $\hat{\rho}$ such that (4.8) holds for all $\rho \ge \hat{\rho}$ is given by

$$\hat{\rho} = \frac{2\|\xi\|}{\|c - s\|}. \tag{4.9}$$

∎

The penalty parameter $\rho$ in NPSQP is computed by maintaining a "current" value of $\rho$ and modifying it if necessary to satisfy (4.2). Thus, the "new" penalty parameter $\bar{\rho}$ is defined by

$$\bar{\rho} = \begin{cases} \rho & \text{if (4.2) holds;} \\ \max(\hat{\rho}, 2\rho) & \text{otherwise,} \end{cases} \tag{4.10}$$

where $\hat{\rho}$ is defined by Lemma 4.3. *Note that $\rho$ can become unbounded only if $\hat{\rho}$ is unbounded.* Furthermore, (4.10) implies that if $\rho$ is bounded over an infinite sequence of iterations, it must eventually reach a finite value, and will retain that value for all subequent iterations.

We now show that, even if the penalty parameter is unbounded, $\rho\|p\|^2$ remains bounded.

**Lemma 4.4.** *If (4.8) holds, then*

$$\hat{\rho}\,\|p\|^2 \le \frac{8\|\xi\|^2}{\gamma}, \tag{4.11}$$

*where $\gamma$ is a lower bound on the smallest eigenvalue of $H$ (cf. assumption (iii)).*

**Proof.** From (4.8) and norm inequalities, we have

$$\tfrac{1}{2}p^THp \le -2(c-s)^T\xi \le 2\|c-s\|\|\xi\|, \tag{4.12}$$

which implies

$$\|c-s\| \ge \frac{\frac{1}{4}p^THp}{\|\xi\|}. \tag{4.13}$$

Combining (4.9) and (4.13) and noting that $p^THp \ge \gamma\|p\|^2$, we obtain

$$\hat{\rho} \le \frac{2\|\xi\|^2}{\frac{1}{4}p^THp} \le \frac{8\|\xi\|^2}{\gamma\|p\|^2}.$$

The desired result follows immediately. ∎

The next two lemmas establish the existence of a step bounded away from zero, *independent of $k$ and the size of $\rho$*, for which a sufficient decrease condition is satisfied.

**Lemma 4.5.** *For $0 \le \alpha \le 1$,*

$$\phi''(\alpha) = -\phi'(0) + q^T\mu + N\|p\|^2,$$

*where $|N|$ is bounded and independent of $k$.*

**Proof.** Using (4.3), we have

$$\nabla^2 L = \begin{pmatrix} \nabla^2 F - \sum\big(\lambda_i + \rho(c_i - s_i)\big)\nabla^2 c_i + \rho A^TA & -A^T & -\rho A^T \\ -A & 0 & I \\ -\rho A & I & \rho I \end{pmatrix},$$

so that

$$\begin{aligned}\phi''(\alpha) = y^T\nabla^2 L(v+\alpha y)y = p^TW(\alpha)p &- \sum \rho\left(c_i(\alpha) - s_i(\alpha)\right)p^T\nabla^2 c_i(\alpha)p \\ &+ \rho\big(A(\alpha)p - q\big)^T\big(A(\alpha)p - q\big) - 2\xi^T\big(A(\alpha)p - q\big),\end{aligned} \tag{4.14}$$

where

$$W(\alpha) = \nabla^2 F(\alpha) - \sum(\lambda_i + \alpha\xi_i)\nabla^2 c_i(\alpha).$$

We now derive bounds on the first two terms on the right-hand side of (4.14). The first term is bounded in magnitude by a constant multiple of $\|p\|^2$ because of assumption (ii) and the

boundedness of $\|\lambda\|$ (from Lemma 4.2). For the second term, we expand $c_i$ in a Taylor series about $x$:

$$c_i(x+\alpha p) = c_i(x) + \alpha a_i(x)^T p + \tfrac{1}{2}\alpha^2 p^T \nabla^2 c_i(x+\theta_i p)p, \tag{4.15}$$

where $0 < \theta_i < \alpha$. Since $s_i(\alpha) = s_i + \alpha q_i$, using (2.9$a$) and multiplying by $\rho$, we have

$$\rho\left(c_i(x+\alpha p) - (s_i + \alpha q_i)\right) = \rho(1-\alpha)\left(c_i(x) - s_i\right) + \rho\,\tfrac{1}{2}\alpha^2 p^T \nabla^2 c_i(x+\theta_i p)p. \tag{4.16}$$

The term $|\rho(c_i(x) - s_i)|$ is uniformly bounded; this follows trivially from the boundedness of $|c_i|$ and $|s_i|$ if $\rho$ is bounded, and from (4.9) if $\rho$ is unbounded. Because $\|\nabla^2 c_i\|$ is uniformly bounded in the region of interest, a similar argument (using Lemma 4.4 if $\rho$ is unbounded) shows that $|\rho\, p^T \nabla^2 c_i(x+\theta_i p)p|$ is uniformly bounded. Therefore,

$$|\rho(c_i(\alpha) - s_i(\alpha))| \le J_i, \tag{4.17}$$

where $J_i$ is bounded and independent of the iteration. Using (4.17), we obtain the overall bound

$$\sum |\rho\left(c_i(\alpha) - s_i(\alpha)\right) p^T \nabla^2 c_i(\alpha) p| \le J\|p\|^2, \tag{4.18}$$

where $J$ is bounded and independent of the iteration.

Now we examine the third term on the right-hand side of (4.14). Using Taylor series, we have

$$a_i(x+\alpha p)^T p = a_i^T p + \alpha p^T \nabla^2 c_i(\bar{\theta}_i) p, \tag{4.19}$$

where $0 < \bar{\theta}_i < \alpha$. From (2.9$b$) and the boundedness of $\|\rho(c-s)\|$ (and Lemma 4.4 if $\rho$ is unbounded), we obtain

$$\rho p^T\left(A(\alpha) - q\right)^T\left(A(\alpha)p - q\right) = \rho(c-s)^T(c-s) + L\|p\|^2, \tag{4.20}$$

where $|L|$ is bounded and independent of the iteration.

Using (4.19) and the boundedness of $\|\xi\|$, the final term on the right-hand side of (4.14) can be written as

$$-2\xi^T\left(A(\alpha)p - q\right) = 2\xi^T\left(c(x) - s\right) + M\|p\|^2, \tag{4.21}$$

where $|M|$ is bounded and independent of the iteration.

Combining (4.20) and (4.21), the last two terms on the right-hand side of (4.14) become

$$\begin{aligned}\rho\left(A(\alpha)p - q\right)^T\left(A(\alpha)p - q\right) - 2\xi^T\left(A(\alpha)p - q\right) &= \rho(c-s)^T(c-s) + 2\xi^T(c-s) + \bar{M}\|p\|^2 \\ &= -\phi'(0) + q^T\mu + \bar{M}\|p\|^2,\end{aligned}$$

where $|\bar{M}|$ is bounded and independent of the iteration (using (4.6) and noting that the largest eigenvalue of $H$ is bounded).

Combining all these bounds gives the required result. ∎

**Lemma 4.6.** *The linesearch in Step 3 of the algorithm defines a step length $\alpha$ $(0 < \alpha \le 1)$ such that*

$$\phi(\alpha) - \phi(0) \le \sigma\alpha\phi'(0), \tag{4.22}$$

*and $\alpha \ge \bar{\alpha}$, where $0 < \sigma < 1$, and $\bar{\alpha} > 0$ is bounded away from zero and independent of the iteration.*

**Proof.** If both conditions (3.1) are satisfied, then $\alpha = 1$ and (4.22) holds with $\alpha$ trivially bounded away from zero.

Assume that (3.1) does not hold (i.e., $\alpha$ is computed by safeguarded cubic interpolation). The existence of a step length $\alpha$ that satisfies conditions (3.2) is guaranteed from standard analysis (see, for example, Moré and Sorensen, 1984). We need to show that $\alpha$ is uniformly bounded away from zero. There are two cases to consider.

First, assume that (3.1*a*) does not hold, i.e., $\phi(1) - \phi(0) > \sigma\phi'(0)$. Since $\phi'(0) < 0$, this implies the existence of at least one positive zero of the function

$$\psi(\alpha) = \phi(\alpha) - \phi(0) - \sigma\alpha\phi'(0).$$

Let $\alpha^*$ denote the smallest such zero. Since $\psi$ vanishes at zero and $\alpha^*$, and $\psi'(0) < 0$, the mean-value theorem implies the existence of a point $\hat{\alpha}$ $(0 < \hat{\alpha} < \alpha^*)$ such that $\psi'(\hat{\alpha}) = 0$, i.e., for which

$$\phi'(\hat{\alpha}) = \sigma\phi'(0).$$

Because $\sigma \le \eta$, it follows that

$$\phi'(\hat{\alpha}) - \eta\phi'(0) = (\sigma - \eta)\phi'(0) \ge 0.$$

Therefore, since the function $\phi'(\alpha) - \eta\phi'(0)$ is negative at $\alpha = 0$, and non-negative at $\hat{\alpha}$, the mean-value theorem again implies the existence of a smallest value $\bar{\alpha}$ $(0 < \bar{\alpha} \le \hat{\alpha})$ such that

$$\phi'(\bar{\alpha}) = \eta\phi'(0). \tag{4.23}$$

The point $\bar{\alpha}$ is the required lower bound on the step length because (4.23) implies that (3.2*b*) will not be satisfied for any $\alpha \in [0, \bar{\alpha})$.

Expanding $\phi'$ in a Taylor series gives

$$\phi'(\bar{\alpha}) = \phi'(0) + \bar{\alpha}\phi''(\theta),$$

where $0 < \theta < \bar{\alpha}$. Therefore, using (4.23) and noting that $\eta < 1$ and $\phi'(0) < 0$, we obtain

$$\bar{\alpha} = \frac{\phi'(\bar{\alpha}) - \phi'(0)}{\phi''(\theta)} = (1 - \eta)\frac{|\phi'(0)|}{\phi''(\theta)}. \tag{4.24}$$

(Since $\bar{\alpha} > 0$, $\theta$ must be such that $\phi''(\theta) > 0$.) We seek a lower bound on $\bar{\alpha}$, and hence an upper bound on the denominator of (4.24). We know from Lemma 4.5 that

$$\phi''(\theta) = -\phi'(0) + q^T\mu + N\|p\|^2,$$

and from (2.10) that $q^T\mu \le 0$. Therefore,

$$\phi''(\theta) \le |\phi'(0)| + |N|\|p\|^2,$$

and hence

$$\bar{\alpha} \ge \frac{(1-\eta)|\phi'(0)|}{|\phi'(0)| + |N|\|p\|^2}.$$

Dividing by $|\phi'(0)|$ gives

$$\bar{\alpha} \ge \frac{(1-\eta)}{1 + \dfrac{|N|\|p\|^2}{|\phi'(0)|}}. \tag{4.25}$$

Since the algorithm guarantees that $\phi'(0) \le -\frac{1}{2}p^T Hp$, it follows that

$$|\phi'(0)| \ge \tfrac{1}{2}p^T Hp \ge \tfrac{1}{2}\gamma\|p\|^2, \tag{4.26}$$

where $\gamma$ is bounded below. Thus, the denominator of (4.25) may be bounded above as follows:

$$1 + \frac{|N|\|p\|^2}{|\phi'(0)|} \le 1 + \frac{|N|\|p\|^2}{\frac{1}{2}\gamma\|p\|^2} = 1 + \frac{2|N|}{\gamma}.$$

A uniform lower bound on $\bar{\alpha}$ is accordingly given by

$$\bar{\alpha} \ge \frac{\gamma(1-\eta)}{\gamma + 2|N|}. \tag{4.27}$$

In the second case, we assume that (3.1*a*) is satisfied, but (3.1*b*) is not. In this case, it must hold that $\phi'(1) \ge 0$, and hence $\phi'(\alpha)$ must have at least one zero in $(0, 1]$. If $\bar{\alpha}$ denotes the smallest of these zeros, $\bar{\alpha}$ satisfies (4.23), and (4.27) is again a uniform lower bound on the step length. ∎

The proof of a global convergence theorem is now straightforward.

**Theorem 4.1.** *Under assumptions (i)–(v), the algorithm defined by (1.2), (1.3), (4.2), and (4.22) has the property that*

$$\lim_{k\to\infty} \|p_k\| = 0. \tag{4.28}$$

**Proof.** If $\|p_k\| = 0$ for any finite $k$, the algorithm terminates and the theorem is true. Hence we assume that $\|p_k\| \ne 0$ for any $k$. There are two cases to consider.

First, assume that $\rho$ is unbounded. In this case, the discussion following Lemma 4.3 shows that $\hat{\rho}$ must be unbounded, and hence we know from Lemma 4.4 that

$$\|p\|^2 \leq \frac{8\|\xi\|}{\gamma\hat{\rho}}. \tag{4.29}$$

Since $\|\xi\|$ is bounded, the desired result follows immediately.

Second, assume that $\rho$ is bounded. The discussion following Lemma 4.3 shows that in this case $\rho$ must eventually become fixed at a given finite value, i.e., there exists a value $\tilde{\rho}$ and an iteration index $\tilde{K}$ such that $\rho = \tilde{\rho}$ for all $k \geq \tilde{K}$. We consider henceforth only such values of $k$.

The proof is by contradiction. We assume that there exists $\epsilon > 0$ and $K \geq \tilde{K}$ such that $\|p_k\| \geq \epsilon$ for $k \geq K$. Every subsequent iteration must therefore yield a strict decrease in the merit function (2.7) with $\rho = \tilde{\rho}$, because, using (4.22),

$$\phi(\alpha) - \phi(0) \leq \eta\alpha\phi'(0) \leq \tfrac{1}{2}\eta\bar{\alpha}\gamma\epsilon^2 < 0.$$

The two final inequalities are derived from Lemma 4.6, since $\alpha \geq \bar{\alpha}$, which is uniformly bounded away from zero. The adjustment of the slack variables $s$ in Step 2 of the algorithm can only lead to a further reduction in the merit function. Therefore, since the merit function with $\rho = \tilde{\rho}$ decreases by at least a fixed quantity at every iteration, it must be unbounded below. But this is impossible, from assumptions (i)–(ii) and Lemma 4.2. Therefore, (4.28) must hold. ∎

**Corollary 4.1.**

$$\lim_{k\to\infty} \|x_k - x^*\| = 0.$$

**Proof.** The result follows immediately from Theorem 4.1 and Lemma 4.1. ∎

The second theorem shows that Algorithm NPSQP also exhibits convergence of the multiplier estimate $\lambda$.

**Theorem 4.2.** *If $\hat{A}(x^*)$, the Jacobian matrix of the constraints active at $x^*$, has full rank, then*

$$\lim_{k\to\infty} \|\lambda_k - \lambda^*\| = 0.$$

**Proof.** If $\|p_k\| = 0$, then $\mu_k = \lambda^*$ (using (1.1) and (1.4)); in Step 1 of Algorithm NPSQP, $\lambda_{k+1}$ is set to $\lambda^*$, and the algorithm terminates. Thus, the theorem is true if $p_k = 0$ for any $k$. We therefore assume that $\|p_k\| \neq 0$ for any $k$.

The definition (4.1) gives

$$\lambda_{k+1} = \sum_{i=0}^{k} \gamma_{ik}\mu_i, \tag{4.30}$$

where

$$\gamma_{kk} = \alpha'_k \quad \text{and} \quad \gamma_{ik} = \alpha'_i \prod_{j=i+1}^{k} (1 - \alpha'_j), \quad i < k, \tag{4.31}$$

with $\alpha'_0 = 1$ and $\alpha'_i = \alpha_i$, $i \geq 1$. (This convention is used because of the special initial condition that $\lambda_0 = \mu_0$.) From Lemma 4.6 and (4.31), we observe that

$$0 < \bar{\alpha} \leq \alpha'_i \leq 1 \quad \text{for all } i, \tag{4.32a}$$

$$\sum_{i=0}^{k} \gamma_{ik} = 1 \tag{4.32b}$$

and

$$\gamma_{ik} \leq (1 - \bar{\alpha})^{k-i}, \quad i < k. \tag{4.32c}$$

Since we know from Theorem 4.1 that $x_k \rightarrow x^*$, the iterates will eventually reach a neighborhood of $x^*$ in which the QP subproblem identifies the correct active set (Lemma 4.1) and $\hat{A}(x_k)$ has full rank. Assume that these properties hold for $k \geq K_1$. From the definition (1.4) of $\mu$, assumptions (ii)–(iii), and the existence of a lower bound on the smallest singular value of $\hat{A}$, we have for $k \geq K_1$ that there exists a bounded scalar $M$ such that

$$\mu_k = \lambda^* + M_k d_k t_k, \tag{4.33}$$

with $|M_k| \leq M$, $d_k = \max(\|p_k\|, \|x^* - x_k\|)$ and $\|t_k\| = 1$. Given any $\epsilon > 0$, Theorem 4.1 and Corollary 4.1 also imply that $K_1$ can be chosen so that, for $k \geq K_1$,

$$|M_k d_k| \leq \tfrac{1}{2}\epsilon. \tag{4.34}$$

We can also define an iteration index $K_2$ with the following property:

$$(1 - \bar{\alpha})^k \leq \frac{\epsilon}{2(k+1)(1 + \bar{\mu} + \|\lambda^*\|)} \tag{4.35}$$

for $k \geq K_2 + 1$, where $\bar{\mu}$ is an upper bound on $\|\mu\|$ for all $k$. Let $K = \max(K_1, K_2)$. Then, from (4.30) and (4.33), we have for $k \geq 2K$,

$$\lambda_{k+1} = \sum_{i=0}^{K} \gamma_{ik}\mu_i + \sum_{i=K+1}^{k} \gamma_{ik}(\lambda^* + M_i d_i t_i).$$

Hence it follows from (4.32b) that:

$$\lambda_{k+1} - \lambda^* = \sum_{i=0}^{K} \gamma_{ik}(\mu_i - \lambda^*) + \sum_{i=K+1}^{k} \gamma_{ik} M_i d_i t_i.$$

From the bounds on $\|\mu_i\|$ and $\|t_i\|$ we then obtain

$$\|\lambda_{k+1} - \lambda^*\| \leq (\bar{\mu} + \|\lambda^*\|) \sum_{i=0}^{K} \gamma_{ik} + \sum_{i=K+1}^{k} \gamma_{ik} |M_i d_i|. \tag{4.36}$$

Since $k \geq 2K$, it follows from (4.32*a*) and (4.32*c*) that

$$\sum_{i=0}^{K} \gamma_{ik} \leq \sum_{i=0}^{K} (1-\bar{\alpha})^{k-i} \leq \sum_{i=0}^{K} (1-\bar{\alpha})^{2K-i} \leq (K+1)(1-\bar{\alpha})^{K}.$$

Using (4.35), we thus obtain the following bound for the first term on the right-hand side of (4.36):

$$(\bar{\mu} + \|\lambda^*\|) \sum_{i=0}^{K} \gamma_{ik} \leq \tfrac{1}{2}\epsilon. \tag{4.37}$$

To bound the second term in (4.36), we use (4.32*b*) and (4.34):

$$\sum_{i=K+1}^{k} \gamma_{ik} |M_i d_i| \leq \tfrac{1}{2}\epsilon \sum_{i=K+1}^{k} \gamma_{ik} \leq \tfrac{1}{2}\epsilon. \tag{4.38}$$

Combining (4.36)-(4.38), we obtain the following result: given any $\epsilon > 0$, we can find $K$ such that

$$\|\lambda_k - \lambda^*\| \leq \epsilon \quad \text{for} \quad k \geq 2K+1,$$

which implies that

$$\lim_{k \to \infty} \|\lambda_k - \lambda^*\| = 0.$$

∎

## 5. Use within a Superlinearly Convergent Algorithm

As mentioned in Section 2.1, a point of interest is whether superlinear convergence may be impeded by the requirement of a sufficient decrease in the merit function at every iteration. In this section we show that a unit step ($\alpha = 1$) will satisfy conditions (3.1) when the iterates are sufficiently close to the solution and $x$ and $\lambda$ are converging superlinearly at the same rate. For further discussion of superlinear convergence, see Dennis and Moré (1977) and Boggs, Tolle and Wang (1982).

In addition to the conditions assumed in the previous section, we assume that for all sufficiently large $k$:

$$x_k + p_k - x^* = o(\|x_k - x^*\|), \tag{5.1a}$$

$$\lambda_k + \xi_k - \lambda^* = o(\|\lambda_k - \lambda^*\|), \tag{5.1b}$$

$$\frac{\|p_k\|}{\|\xi_k\|} > M > 0, \tag{5.1c}$$

where $M$ is independent of $K$. Note that (5.1) implies

$$\|p_k\| \sim \|x_k - x^*\| \quad \text{and} \quad \|\xi_k\| \sim \|\lambda_k - \lambda^*\|, \tag{5.2}$$

where "$\sim$" means that the two quantities are of similar order. (See also Dennis and Moré, 1977.)

First we show that these assumptions imply that the penalty parameter $\rho$ in the merit function (2.7) must remain bounded for all $k$.

**Lemma 5.1.** *Under assumptions (i)–(v) of Section 3, and conditions (5.1), there exists a finite $\bar{\rho}$ such that for all $k$,*

$$\rho \le \bar{\rho}. \tag{5.3}$$

**Proof.** Recall from the proof of Lemma 4.3 that $\rho$ can become unbounded only if (4.8) holds and $\hat{\rho}$ as defined by (4.9) becomes unbounded. Condition (4.8) states that $-2(c-s)^T\xi \ge \frac{1}{2}p^T Hp$. Using assumption (iii) and norm inequalities, (4.8) thus implies

$$\|c-s\| \ge \frac{\frac{1}{4}\gamma\|p\|^2}{\|\xi\|}.$$

Hence, from (5.1c), we obtain

$$\frac{\|c-s\|}{\|\xi\|} \ge \frac{\frac{1}{4}\gamma\|p\|^2}{\|\xi\|^2} = \bar{M} > 0.$$

Consequently, from Lemma 4.4,

$$\hat{\rho} < \frac{2}{\bar{M}},$$

and remains bounded. This proves the lemma. ∎

The next two lemmas show that conditions (3.1) are eventually satisfied for all $k$ sufficiently large.

**Lemma 5.2.** *Under assumptions (i)–(v) of Section 3 and conditions (5.1), the sufficient decrease condition (3.1a) holds for sufficiently large $k$, i.e.,*

$$\phi(1) - \phi(0) \le \sigma\phi'(0),$$

*where $0 < \sigma < \frac{1}{2}$.*

**Proof.** As in Powell and Yuan (1984), observe that the continuity of second derivatives gives the following relationships:

$$F(x+p) = F(x) + \tfrac{1}{2}\big(g(x) + g(x+p)\big)^T p + o(\|p\|^2),$$
$$c(x+p) = c(x) + \tfrac{1}{2}\big(A(x) + A(x+p)\big)p + o(\|p\|^2).$$

Conditions (5.1) and (5.2) then imply:

$$F(x+p) = F(x) + \tfrac{1}{2}\big(g(x) + g(x^*)\big)^T p + o(\|p\|^2), \tag{5.4a}$$

$$c(x+p) = c(x) + \tfrac{1}{2}\big(A(x) + A(x^*)\big)p + o(\|p\|^2). \tag{5.4b}$$

We shall henceforth use $g$ to denote $g(x)$ and $g^*$ to denote $g(x^*)$, and similarly for $F$, $c$ and $A$.

By definition,

$$\phi(0) = F - \lambda^T(c - s) + \tfrac{1}{2}\rho(c - s)^T(c - s), \tag{5.5a}$$

$$\begin{aligned}\phi(1) &= F(x + p) - \mu^T(c(x + p) - c - Ap)\\ &\quad + \tfrac{1}{2}\rho\big(c(x + p) - c - Ap\big)^T\big(c(x + p) - c - Ap\big).\end{aligned} \tag{5.5b}$$

Using Taylor series, we have

$$c(x + p) = c + Ap + O(\|p\|^2). \tag{5.6}$$

Substituting from (5.4) and (5.6) into (5.5$b$), we obtain

$$\phi(1) = F + \tfrac{1}{2}(g + g^*)^T p - \tfrac{1}{2}\mu^T(A - A^*)p + o(\|p\|^2). \tag{5.7}$$

Combining (5.5$a$) and (5.7) gives

$$\begin{aligned}\phi(1) - \phi(0) &= \tfrac{1}{2}p^Tg + \tfrac{1}{2}p^Tg^* - \tfrac{1}{2}\mu^TAp + \tfrac{1}{2}\mu^TA^*p + \lambda^T(c - s)\\ &\quad - \tfrac{1}{2}\rho(c - s)^T(c - s) + o(\|p\|^2).\end{aligned} \tag{5.8}$$

Using (2.6), (2.9) and (4.4), we obtain the following expression:

$$\phi'(0) = p^Tg + 2\lambda^T(c - s) - \rho(c - s)^T(c - s) + \mu^TAp - \mu^Tq. \tag{5.9}$$

Substituting (5.9) into (5.8), we have

$$\phi(1) - \phi(0) = \tfrac{1}{2}\phi'(0) + \tfrac{1}{2}\mu^Tq + \tfrac{1}{2}p^T(g^* - A^{*T}\mu) + o(\|p\|^2). \tag{5.10}$$

It follows from (5.1$b$)–(5.1$c$) that the expression $g^* - A^{*T}\mu$ is $o(\|p\|)$, which gives

$$\begin{aligned}\phi(1) - \phi(0) &= \tfrac{1}{2}\phi'(0) + \tfrac{1}{2}\mu^Tq + o(\|p\|^2)\\ &\le \tfrac{1}{2}\phi'(0) + o(\|p\|^2),\end{aligned}$$

and hence

$$\phi(1) - \phi(0) - \sigma\phi'(0) \le (\tfrac{1}{2} - \sigma)\phi'(0) + o(\|p\|^2). \tag{5.11}$$

Since $\sigma < \frac{1}{2}$ and $\phi'(0)$ satisfies (4.2), (5.11) implies that (3.1$a$) holds for sufficiently large $k$. ∎

**Lemma 5.3.** *Under assumptions (i)–(v) and conditions (5.1), the second linesearch condition (3.1b) holds for sufficiently large $k$, i.e.,*

$$\phi'(1) \le \eta\phi'(0) \quad \text{or} \quad |\phi'(1)| \le \eta|\phi'(0)|,$$

*where $\eta < \frac{1}{2}$.*

**Proof.** In this proof, we use the notation $g(1)$ to denote $g(x + p)$, and similarly for $c$ and $A$. Using (4.3), $\phi'(1)$ is given by

$$\begin{aligned}\phi'(1) &= p^Tg(1) - p^TA(1)\mu + \rho p^T(c(1) - c - Ap)\\ &\quad + (\mu - \lambda)^T\big(c(1) - c - Ap\big) + q^T\mu - \rho q^T\big(c(1) - c - Ap\big).\end{aligned} \tag{5.12}$$

From conditions (5.2)–(5.3), we have

$$g(1) = g^* + o(\|p\|^2), \quad A(1) = A^* + o(\|p\|^2) \quad \text{and} \quad \mu = \lambda^* + o(\|p\|). \tag{5.13}$$

Substituting from (5.6) and (5.13) into (5.12) then gives:

$$\phi'(1) = p^T g^* - p^T A^* \mu + q^T \mu - \rho(c - s)^T \big(c(1) - c - Ap\big) + o(\|p\|^2). \tag{5.14}$$

Consider now the vectors $(c - s)$, $q$ and $\mu$. From (2.8) and (5.2), $\|c - s\| = O(\|p\|)$. From Lemma 4.1, we know that the QP subproblem (1.3) will eventually predict the correct active set, and hence that $\mu_i = 0$ if $c_i(x^*) > 0$. For an active constraint $c_i$, it follows from (2.8) that $s_i$ will eventually be set to zero at the beginning of every iteration if $\rho > 0$, and hence $q_i$ must also be zero. If $\rho = 0$ and $c_i$ is an active constraint, then $|q_i| = o(\|p\|^2)$. Therefore, in either case we have that

$$|q^T \mu| = o(\|p\|^2) \quad \text{and} \quad |\rho(c - s)^T \big(c(1) - c - Ap\big)| = o(\|p\|^2). \tag{5.15}$$

Recalling that $\|g^* - A^{*T}\mu\| = o(\|p\|)$ and using (5.15) in (5.14), we obtain

$$\phi'(1) = o(\|p\|^2). \tag{5.16}$$

Since $|\phi'(0)| \geq \frac{1}{2}\gamma\|p\|^2$, (5.16) imples that (3.1$b$) will eventually be satisfied at every iteration. ∎

## 6. Numerical Results

In order to indicate the reliability and efficiency of a practical SQP algorithm based on the merit function (2.7), we present a selection of numerical results obtained from the Fortran code NPSOL (Gill *et al.*, 1986a, b). Table 1 contains the results of solving a subset of problems 70–119 from Hock and Schittkowski (1981). (We have omitted problems that are non-differentiable or that contain only linear constraints. Since NPSOL treats linear constraints separately, they do not affect the merit function.)

The problems were run on an IBM 3081K, in double precision (i.e., machine precision $\epsilon_M$ is approximately $2.22 \times 10^{-16}$). The default parameters for NPSOL were used in all cases (for details, see Gill *et al.*, 1986a). In particular, the default value for *ftol*, the final accuracy requested in $F$, is $5.4 \times 10^{-12}$, and *ctol*, the feasibility tolerance, is $\sqrt{\epsilon_M}$. Analytic gradients were provided for all functions, and a gradient linesearch was used.

For successful termination of NPSOL, the iterative sequence of $x$-values must have converged and the final point must satisfy the first-order Kuhn-Tucker conditions (*cf.* (1.1)). The sequence of iterates is considered to have converged at $x$ if

$$\alpha\|p\| \leq \sqrt{ftol}\,(1 + \|x\|), \tag{6.1}$$

where $p$ is the search direction and $\alpha$ the step length from (1.2). The iterate $x$ is considered to satisfy the first-order conditions for a minimum if the following conditions hold: the inactive

constraints are satisfied to within *ctol*, the magnitude of each active constraint residual is less than *ctol*, and

$$\|Z(x)^T g(x)\| \le \sqrt{ftol}\,\big(1 + \max(1 + |F(x)|, \|g(x)\|)\big), \tag{6.2}$$

where $Z(x)$ is an orthogonal basis for the null space of the gradients of the active constraints. (Thus, $Z(x)^T g(x)$ is the usual *projected gradient*.)

Table 1 gives the following information: the problem number in Hock and Schittkowski; the number of variables ($n$); the number of simple bounds ($m_B$); the number of general linear constraints ($m_L$); the number of nonlinear constraints ($m_N$); the number of iterations; and the number of function evaluations. (Each constraint is assumed to include a lower and an upper bound.)

**Table 1**

| Problem | $n$ | $m_B$ | $m_L$ | $m_N$ | Iterations | Evaluations |
|---|---|---|---|---|---|---|
| 70 | 4 | 4 | 0 | 1 | 35 | 38 |
| 71 | 4 | 4 | 0 | 2 | 5 | 6 |
| 72 | 4 | 4 | 0 | 2 | 6 | 7 |
| 73 | 4 | 4 | 2 | 1 | 3 | 4 |
| 74 | 4 | 4 | 2 | 3 | 9 | 10 |
| 75 | 4 | 4 | 2 | 3 | 7 | 8 |
| 77 | 5 | 0 | 0 | 2 | 14 | 20 |
| 78 | 5 | 0 | 0 | 3 | 8 | 10 |
| 79 | 5 | 0 | 0 | 3 | 9 | 12 |
| 80 | 5 | 5 | 0 | 3 | 8 | 10 |
| 81 | 5 | 5 | 0 | 3 | 9 | 10 |
| 83 | 5 | 5 | 0 | 3 | 4 | 6 |
| 84 | 5 | 5 | 0 | 3 | 5 | 6 |
| 85 | 5 | 5 | 0 | 38 | 45 | 68 |
| 93 | 6 | 6 | 0 | 2 | 11 | 14 |
| 95 | 6 | 6 | 0 | 4 | 1 | 2 |
| 96 | 6 | 6 | 0 | 4 | 1 | 2 |
| 97 | 6 | 6 | 0 | 4 | 3 | 6 |
| 98 | 6 | 6 | 0 | 4 | 3 | 6 |
| 99 | 7 | 7 | 0 | 2 | 19 | 34 |
| 100 | 7 | 0 | 0 | 4 | 15 | 34 |
| 101 | 7 | 7 | 0 | 5 | 33 | 85 |
| 102 | 7 | 7 | 0 | 5 | 29 | 69 |
| 103 | 7 | 7 | 0 | 5 | 25 | 60 |
| 104 | 8 | 8 | 0 | 5 | 17 | 19 |
| 106 | 8 | 8 | 3 | 3 | 43 | 47 |
| 107 | 9 | 7 | 0 | 6 | 11 | 18 |
| 108 | 9 | 1 | 0 | 13 | 15 | 19 |
| 109 | 9 | 9 | 1 | 8 | 18 | 19 |
| 111 | 10 | 10 | 0 | 3 | 59 | 62 |
| 113 | 10 | 0 | 3 | 5 | 14 | 19 |
| 114 | 10 | 10 | 5 | 6 | 51 | 81 |
| 116 | 13 | 13 | 5 | 10 | 54 | 78 |
| 117 | 15 | 15 | 0 | 5 | 19 | 20 |

NPSOL terminated successfully on all the problems except the badly scaled problem 85, for which (6.1) could not be satisfied. However, the optimality conditions were satisfied at the final iterate, which gave an improved objective value compared to that in Hock and Schittkowski (1981).

## 7. References

Bartholomew-Biggs, M. C. (1983). A recursive quadratic programming algorithm based on the augmented Lagrangian function, Report 139, Numerical Optimisation Centre, Hatfield Polytechnic, Hatfield, England.

Bartholomew-Biggs, M. C. (1985). The development of recursive quadratic programming methods based on the augmented Lagrangian, Report 160, Numerical Optimisation Centre, Hatfield Polytechnic, Hatfield, England.

Boggs, P. T. and Tolle, J. W. (1984). A family of descent functions for constrained optimization, *SIAM J. on Numerical Analysis* 21, 1146–1161.

Boggs, P. T. and Tolle, J. W. (1985). An efficient strategy for utilizing a merit function in nonlinear programming algorithms, Report 85-5, Department of Operations Research and Systems Analysis, University of North Carolina, North Carolina.

Boggs, P. T., Tolle, J. W. and Wang, P. (1982). On the local convergence of quasi-Newton methods for constrained optimization, *SIAM J. on Numerical Analysis* 161–171.

Chamberlain, R. W., Lemaréchal, C., Pedersen, H., and Powell, M. J. D. (1982). The watchdog technique for forcing convergence in algorithms for constrained optimization, *Mathematical Programming Study* 16, 1–17.

Dennis, J. E., Jr. and Moré, J. J. (1977). Quasi-Newton methods, motivation and theory, *SIAM Review* 19, 46–89.

DiPillo, G. and Grippo, L. (1979). A new class of augmented Lagrangians in nonlinear programming, *SIAM J. on Control and Optimization* 17, 618–628.

Dixon, L. C. W. (1979). Exact penalty functions in nonlinear programming, Report 103, Numerical Optimisation Centre, Hatfield Polytechnic, Hatfield, England.

Fiacco, A. V. and McCormick, G. P. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York and Toronto.

Fletcher, R. (1970). A class of methods for nonlinear programming with termination and convergence properties, in J. Abadie (ed.), *Integer and Nonlinear Programming*, North-Holland, Amsterdam, 157–175.

Fletcher, R. (1974). Methods related to Lagrangian functions, in P. E. Gill and W. Murray (eds.), *Numerical Methods for Constrained Optimization*, Academic Press, London and New York, 219–240.

Gill, P. E. and Murray, W. (1974). Safeguarded steplength algorithms for optimization using descent methods, Report NAC 37, National Physical Laboratory, England.

Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H. (1985). Model building and practical aspects of nonlinear programming, in K. Schittkowski (ed.), *Computational Mathematical Programming*, NATO ASI Series F: Computer and Systems Sciences 15, Springer-Verlag, Berlin and New York, 209–247.

Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H. (1986a). User's Guide for NPSOL (Version 4.0): a Fortran package for nonlinear programming, Report SOL 86-2, Department of Operations Research, Stanford University.

Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H. (1986b). A sequential quadratic programming method for nonlinear programming, to appear.

Gill, P. E., Murray, W. and Wright, M. H. (1981). *Practical Optimization*, Academic Press, London and New York.

Han, S.-P. (1976). Superlinearly convergent variable metric algorithms for general nonlinear programming problems, *Mathematical Programming* 11, 263–282.

Han, S.-P. (1977). A globally convergent method for nonlinear programming, *J. Optimization Theory and Applications* 22, 297–310.

Hestenes, M. R. (1969). Multiplier and gradient methods, *J. Optimization Theory and Applications* 4, 303–320.

Hock, W. and Schittkowski, K. (1981). *Test Examples for Nonlinear Programming*, Lecture Notes in Economics and Mathematical Systems, Volume 187, Springer-Verlag, Berlin, Heidelberg and New York.

Maratos, N. (1978). *Exact Penalty Function Algorithms for Finite-Dimensional and Control Optimization Problems*, Ph. D. Thesis, University of London.

Moré, J. J. and Sorensen, D. C. (1984). Newton's method, in G. H. Golub (ed.), *Studies in Numerical Analysis*, Mathematical Association of America, 29–82.

Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, London and New York.

Pietrzykowski, T. (1969). An exact potential method for constrained maxima, *SIAM J. on Numerical Analysis* 6, 299–304.

Powell, M. J. D. (1969). A method for nonlinear constraints in minimization problems, in R. Fletcher (ed.), *Optimization*, Academic Press, London and New York, 283–298.

Powell, M. J. D. (1974). Introduction to constrained optimization, in P. E. Gill and W. Murray (eds.), *Numerical Methods for Constrained Optimization*, Academic Press, London and New York, 1–28.

Powell, M. J. D. (1977). A fast algorithm for nonlinearly constrained optimization calculations, Report DAMTP 77/NA 2, University of Cambridge, England.

Powell, M. J. D. (1983). Variable metric methods for constrained optimization, in A. Bachem, M. Grötschel and B. Korte (eds.), *Mathematical Programming: The State of the Art*, Springer-Verlag, Berlin and New York, 288–311.

Powell, M. J. D. and Yuan, Y. (1984). A recursive quadratic programming algorithm that uses differentiable exact penalty functions, Report DAMTP 84/NA9, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England.

Robinson, S. M. (1974). Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms, *Mathematical Programming* 7, 1–16.

Rockafellar, R. T. (1973). The multiplier method of Hestenes and Powell applied to convex programming, *J. Optimization Theory and Applications* 12, 555–562.

Schittkowski, K. (1981). The nonlinear programming method of Wilson, Han and Powell with an augmented Lagrangian type line search function, *Numerische Mathematik* 38, 83–114.

Schittkowski, K. (1983). On the convergence of a sequential quadratic programming method with an augmented Lagrangian line search function, *Math. Operationsforschung u. Statistik, Ser. Optimization* 14, 197–216.

Tapia, R. A. (1977). Diagonalized multiplier methods and quasi-Newton methods for constrained optimization, *J. Optimization Theory and Applications* 22, 135–194.

Wright, M. H. (1976). *Numerical Methods for Nonlinearly Constrained Optimization*, Ph.D. Thesis, Stanford University, California.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>SOL 86-6 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Some Theoretical Properties of an Augmented Lagrangian Merit Function | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical Report |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Philip E. Gill, Walter Murray, Michael A. Saunders & Margaret H. Wright | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-85-K-0343<br>DAAG29-84-K-0156 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Department of Operations Research - SOL<br>Stanford University<br>Stanford, CA 94305-4022 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>NR-047-064 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research - Dept. of the Navy<br>800 N. Quincy Street<br>Arlington, VA 22217<br><br>U.S. Army Research Office<br>P.O. Box 12211<br>Research Triangle Park, NC 27709 | | 12. REPORT DATE<br>April 1986 |
| | | 13. NUMBER OF PAGES<br>22 pp. |
| | | 15. SECURITY CLASS. (of this report)<br>UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

SQP methods
nonlinear constraints
merit function
augmented Lagrangian
global convergence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See next page.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SOL 86-6: "Some Theoretical Properties of an Augmented Lagrangian Merit Function," Philip E. Gill, Walter Murray, Michael A. Saunders & Margaret H. Wright

Sequential quadratic programming (SQP) methods for nonlinearly constrained optimization typically use a merit function to enforce convergence from an arbitrary starting point. We define a smooth augmented Lagrangian merit function in which the Lagrange multiplier estimate is treated as a separate variable, and inequality constraints are handled by means of non-negative slack variables that are included in the linesearch. Global convergence is proved for an SQP algorithm that uses his merit function. We also prove that steps of unity are accepted in a neighborhood of the solution when this merit function is used in a suitable superlinearly convergent algorithm. Finally, a selection of numerical results is presented to illustrate the performance of the associated SQP method.

END

DTIC

7-86